OPERATIONS RESEARCH
Vol. 58, No. 3, May–June 2010, pp. 756–767
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5803 | 0756

DOI 10.1287/opre.1090.0738

Series Expansions for Continuous-Time Markov Processes

Bernd Heidergott
Department of Econometrics, Vrije Universiteit Amsterdam, and Tinbergen Institute,
1081 HV Amsterdam, The Netherlands, bheidergott@feweb.vu.nl

Arie Hordijk
Mathematical Institute, Leiden University, 2300 RA Leiden, The Netherlands,
hordijk@math.leidenuniv.nl

Nicole Leder
Department of Mathematics, University of Hamburg, 20146 Hamburg, Germany,
nicole.leder@studium.uni-hamburg.de

We present update formulas that allow us to express the stationary distribution of a continuous-time Markov process with denumerable state space having generator matrix Q^* through a continuous-time Markov process with generator matrix Q. Under suitable stability conditions, numerical approximations can be derived from the update formulas, and we show that the algorithms converge at a geometric rate. Applications to sensitivity analysis and bounds on perturbations are discussed as well. Numerical examples are presented to illustrate the efficiency of the proposed algorithm.

Subject classifications: probability: Markov processes; queues: algorithms; numerical methods; retrial queue; deviation matrix; sensitivity analysis.
Area of review: Stochastic Models.
History: Received August 2007; revisions received July 2008, November 2008; accepted March 2009. Published online in *Articles in Advance* October 28, 2009.

1. Introduction

Markov processes are a well-established modeling tool in operations research. In analyzing the long-run behavior of a Markovian system, its stationary distribution plays a key role because, provided the stationary distribution is known, all kinds of long-run performance characteristics can be deduced from it. In many systems that are of interest in applications, the stationary distribution cannot be obtained in a closed analytical form. An important class of systems for which this is the case are retrial queues, which are frequently applied in the modeling and analysis of call centers. For overviews on queueing models for call centers, see, for example, Gans et al. (2003), and Koole and Mandelbaum (2002). Here, even for finite state retrial queues, the stationary distribution typically cannot be obtained in a closed form; see, for example, Falin and Templeton (1997), Kulkarni and Liang (1997), Yang and Templeton (1987), and Falin (1990).

This paper presents an approach to approximately compute the stationary distribution of a "difficult" system, such as a retrial queue, by characteristics of a simpler one. In particular, our approach does not require the Markov processes to be uniformizable, which is a condition typically imposed in the literature in analyzing infinite continuous-time processes and which rules out such simple systems as the M/M/∞ queue. More specifically, we establish a sufficient condition for the fast convergence of our approximation algorithm without requiring uniformization. Subsequently, alternative (and easy to check) conditions will be established under the assumption of uniformizability.

Let $\mathscr{X} = \{X_t,\ t \geqslant 0\}$ be a continuous-time ergodic Markov process on a denumerable state space S describing the nominal system. Throughout this paper we will denote its transition matrix by $P(t)$, and its infinitesimal generator by Q. We assume that $\mathscr{X}$ has a unique stationary distribution, denoted by π. Suppose that Q^* is the generator matrix of another continuous-time Markov process $\mathscr{X}^* = \{X_t^*,\ t \geqslant 0\}$ defined on the same state space as $\mathscr{X}$ and denote its stationary distribution by π^* (existence assumed). In this paper we will present a general update formula that allows us to express π^* as a mapping of Q^*, Q, and π only. More precisely, we will show that

$$\pi^* = \pi \sum_{n\geqslant 0} ((Q^* - Q)D)^n, \tag{1}$$

where $D = (d_{ij})_{i,\, j\in S}$ is the deviation matrix associated with Q, which has elements

$$d_{ij} \stackrel{\text{def}}{=} \int_0^\infty (p_{ij}(t) - \pi_j)\, dt, \quad i, j \in S. \tag{2}$$

The deviation matrix exists whenever all integrals in (2) are finite. Hence, provided the deviation matrix exists and the update formula converges, π^* can be presented as a power series in Q^*, for given Q, D, and π. As a first application of the update formula, we will derive an expression for derivatives of π with respect to parameters of Q. Also, bounds on perturbations will be established. This type of perturbation analysis has a long tradition in Markov chain theory. Early references for discrete-time Markov chains are Schweitzer (1968), and Takahashi (1973). An exhaustive literature review will be given later in the text.

Based on (1), we derive a numerical algorithm for finite state space Markov processes that allows us to compute π^* out of Q^*. In applications this has the nice feature that once D and π are evaluated (either by an explicit formula or numerically), our algorithm offers the opportunity of computing π^* in a fast way. In particular, we will show that the algorithm converges exponentially fast, and we will provide an efficient bound on the error made by evaluating the update formula only up to a finite number of elements. This provides a tool for discrete optimization because, once π and D are computed, π^* can be computed for various choices of Q^* by simple matrix multiplication, and we consider the update formula as a fast and efficient way of simultaneously analyzing the effect of changing Q to Q^* on the stationary distribution for various possible alternatives.

In addition, a functional version of the algorithm exists that applies when π is a function of θ where θ is a parameter of Q, which is expressed in the notation by writing Q_θ. Denoting the stationary distribution of Q_θ by π_θ, we develop $\pi_{\theta+\Delta}$ as a polynomial in Δ through a *functional series expansion*. This provides a tool for numerical analysis in the following way. Suppose that one is interested in the maximal value of θ such that $\pi_\theta g \leqslant c$ for some cost function g and given constant c. Then, the functional series expansion can be used for obtaining $\pi_{\theta+\Delta}g$ as a polynomial in Δ. Hence, if $\pi_{\theta+\Delta}g = c$ can be solved for, say Δ^*, then $\theta^* = \theta + \Delta^*$ yields an (approximate) solution to the problem $\max_\theta \pi_\theta g$ subject to $\pi_\theta g \leqslant c$.

The paper is organized as follows. Section 2 provides an overview on the literature and discusses the relationship of our approach to the state of the art in the literature. In §3 basic properties of continuous-time Markov processes are presented. The update formula and its application to sensitivity analysis and bounding of perturbations is investigated in §4. Numerical algorithms are discussed in §5. Functional series expansions are presented in §6. The discussion of the algorithms in §5 and §6 is supported by retrial queue examples. Proofs of the more technical results are provided in an electronic companion available as part of the online version that can be found at http://or.journal.informs.org/.

2. Literature Review

There exists an extensive literature on perturbation analysis of Markov chains (PAMC). This research was initiated by the papers Schweitzer (1968), and Takahashi (1973), which studied finite state-space discrete-time Markov chains. The study of PAMC for continuous-time processes started somewhat later, and, to the best of our knowledge, the first reference on this is Tweedie (1980). In this paper it is assumed that the generator matrix Q of a continuous-time denumerable state space Markov process is perturbed by "small perturbations" so that the perturbed generator Q^* satisfies

$$\frac{1}{1+\epsilon_i} Q_{ij} \leqslant Q^*_{ij} \leqslant (1+\epsilon_i) Q_{ij}, \quad i \neq j,$$

provided that

$$(1+\epsilon) = \prod_i (1+\epsilon_i) < \infty.$$

Then, it holds for the stationary distribution associated with Q^* that

$$\left(\frac{1}{1+\epsilon}\right)^k \pi_i \leqslant \pi^*_i \leqslant (1+\epsilon)^k \pi_i,$$

where $k = 4$ for general processes and $k = 2$ in case that Q is uniformizable.

There exists a considerable literature on perturbation analysis of continuous-time Markov processes with finite state space. For references, see Mitrophanov (2004; 2005a, b; 2006), Mitrophanov and Borodovsky (2006), Mitrophanov et al. (2005), and Yin et al. (2002).

For continuous-time Markov processes with denumerable state space, PAMC is usually carried out under the assumption of uniformizablility. Under this assumption and together with the additional assumption of strong ergodicity (a Markov chain is called strongly ergodic if it is geometrically ergodic with respect to the total variational norm), perturbation bounds and differentiation formulas are established in Cao and Chen (1997). Again under uniformizability, Altman et al. (2004) established series expansions for π^* in terms of π, Q, and Q^*. Moreover, this paper also considered so-called singular perturbations, where a perturbation is called singular if Q has several ergodic classes whereas Q^* has one unique ergodic class.

Dropping uniformizability, special types of perturbations have been addressed in the literature. One line of research is the study of similar Markov processes, where it is assumed that $Q^*_{ij} = c_{ij} Q_{ij}$ for positive constants c_{ij}; see Pollett (2001), Lenin et al. (2000), and Di Crescenzo (1994a, b). Another line of research is studying the effect of modeling a continuous-time process by the embedded chain at time instances that are multiples of Δt. This can be found in Shardlow and Stuart (2000). For a result on the perturbation of only finitely many elements of Q, see Leskelä (2006).

The approach presented in this paper overcomes the restrictions of the aforementioned papers because it neither requires uniformizability of Q or strong stability, nor does it assume any special structure of the perturbations other than that they are regular, i.e., the perturbed chain is also a unichain.

3. Preliminaries on Denumerable Markov Processes

Throughout this paper we assume that infinitesimal generators are *conservative*, where a matrix is called conservative if its row-sums are equal to zero. Moreover, we assume that Markov processes are *irreducible* and *positive recurrent*, where a Markov process is called irreducible and positive recurrent if from every state $i \in S$ all states $j \in S$ are accessible and the first return time for every state is finite. For the stationary distribution π we denote the associated *ergodic matrix* by Π, where Π is a matrix with rows equal to π.

3.1. Basic Properties of the Generator Matrix

We summarize basic properties of the infinitesimal generator Q, the ergodic matrix Π, and the deviation matrix D.

Lemma 3.1. *If it exists, the ergodic matrix of a continuous-time Markov process $\mathscr{X}$ with infinitesimal generator Q satisfies*

(i) $A\Pi = 0$ *for all conservative matrices* $A \in \mathbb{R}^{S\times S}$, *and even* $\Pi Q = 0$; *and*

(ii) $B\Pi = \Pi$ *for all stochastic matrices* $B \in \mathbb{R}^{S\times S}$.

If, in addition, the deviation matrix exists, then it holds that

(iii) $\Pi D = 0$; *and*

(iv) $-QD = I - \Pi$.

Definition 3.2. A generator matrix Q is called uniformizable with rate μ if $\mu = \sup_j |q_{jj}| < \infty$.

Although any finite-dimensional generator matrix is uniformizable, a classical example of a Markov process on denumerable state space that fails to have this property is the M/M/∞ queue. Note that if Q is uniformizable with rate μ, then Q is uniformizable with rate η for any $\eta > \mu$.

Let Q be uniformizable with rate μ and introduce the Markov chain P_μ as follows:

$$[P_\mu]_{ij} = \begin{cases} q_{ij}/\mu & i \neq j \\ 1 + q_{ii}/\mu & i = j, \end{cases} \tag{3}$$

for i, $j \in S$, or, in shorthand notation,

$$P_\mu = I + \frac{1}{\mu} Q,$$

then it holds that

$$P(t) = e^{-\mu t} \sum_{n=0}^{\infty} \frac{(\mu t)^n}{n!} (P_\mu)^n, \quad t \geqslant 0. \tag{4}$$

Moreover, the stationary distribution of P_μ and $P(t)$ coincide, in formula: $\Pi_\mu = \Pi$. The Markov chain $\mathscr{X}_\mu = \{X_n^\mu : n \geqslant 0\}$ with transition probability matrix P_μ is called the *sampled chain*. The relationship between $\mathscr{X}$ and $\mathscr{X}_\mu$ can be expressed as follows. Let $N_\mu(t)$ denote a Poisson process with rate μ, then $X^\mu_{N_\mu(t)}$ and X_t are equal in distribution for all $t \geqslant 0$. The deviation matrix associated with P_μ is defined by

$$D_\mu = \sum_{n\geqslant 0} ((P_\mu)^n - \Pi_\mu) = \sum_{n\geqslant 0} (P_\mu - \Pi_\mu)^n - \Pi_\mu,$$

provided the sum exists. If Q is uniformizable with rate μ and the deviation matrices associated with P_μ and $P(t)$ exist, then

$$\frac{1}{\mu} D_\mu = D; \tag{5}$$

for a proof, see Coolen-Schrijner and van Doorn (2002). Let Q^* be uniformizable with rate μ^* and let Q be uniformizable with rate μ. Let $\eta = \max(\mu, \mu^*)$, then (3) implies that

$$P_\eta^* - P_\eta = \frac{1}{\eta}(Q^* - Q)$$

and by (5) (with $\mu = \eta$) it follows that

$$(P_\eta^* - P_\eta) D_\eta = (Q^* - Q) D. \tag{6}$$

3.2. Geometric Ergodicity

The main tool for our analysis is the weighted supremum norm, also called v-norm, denoted by $\|\cdot\|_v$, where v is some vector with elements $v_i \geqslant 1$ for all $i \in S$, and for any $w \in \mathbb{R}^{\mathscr{S}}$

$$\|w\|_v \stackrel{\text{def}}{=} \sup_{i\in\mathscr{S}} \frac{|w(i)|}{v(i)}. \tag{7}$$

For a matrix $A \in \mathbb{R}^{S\times S}$ the v-norm is given by

$$\|A\|_v \stackrel{\text{def}}{=} \sup_{i,\, \|w\|_v \leqslant 1} \frac{\sum_{j=1}^{S} |A(i,j) w(j)|}{v(i)},$$

which implies

$$\max_{j\in S} |A(i,j)| \leqslant \|A\|_v v(i), \quad i \in S. \tag{8}$$

Note that v-norm convergence to 0 implies elementwise convergence to 0. With the help of the above concepts, v-geometric ergodicity (also called v-normed ergodicity) of $P(t)$ can be introduced as follows.

Definition 3.3. The Markov process $\mathscr{X}$ is v-geometric ergodic if $c < \infty$ and $\beta < 1$ exist such that

$$\|P(t) - \Pi\|_v \leqslant c\beta^t$$

for all $t \geqslant 0$.

Note that

$$\|D\|_v \leqslant \int_0^\infty \|P(t)-\Pi\|_v \, dt,$$

and it is straightforward to check that v-geometric ergodicity implies existence of $\|D\|_v$ by assuring the finiteness of its elements. Unfortunately, v-geometric ergodicity is almost impossible to check in a direct way. One of the reasons is that $P(t)$ is in general not known in explicit form. If Q is uniformizable, v-norm ergodicity of $P(t)$ can be deduced from v-normed ergodicity of P_μ in discrete time.

In the following we summarize important properties of uniformizable processes.

LEMMA 3.4. *Let Q be uniformizable with rate μ.*

(i) *If finite constants c and β, with $0 \leqslant \beta < 1$, exist, such that $\|(P_\mu)^n - \Pi\|_v \leqslant c\beta^n$ for all n, then $P(t)$ is v-norm ergodic.*

(ii) *If P_μ is v-norm ergodic, then $\|D\|_v$ is finite.*

(iii) *If P_μ is v-norm ergodic, then D_μ and D exist and it holds that*

$$D_\mu = \sum_{n\geqslant 0}((P_\mu)^n - \Pi_\mu) \quad \textit{and} \quad D = \frac{1}{\mu}\sum_{n\geqslant 0}((P_\mu)^n - \Pi).$$

Let

$$m_{e,j} = \sum_{i\in S} \pi_i m_{i,j},$$

where $m_{i,j}$ is the mean first passage time from state i to state j. Kemeny and Snell (1960) showed that $D = \sum_{n\geqslant 0}(I+Q-\Pi)^n - \Pi$ provided that $m_{e,j} < \infty$ for all $j \in S$. According to Coolen-Schrijner and van Doorn (2002), the aforementioned condition on $m_{e,j}$ can be relaxed to the following: There is one $j \in S$ such that $m_{e,j} < \infty$. The precise statement is given in the lemma below.

LEMMA 3.5. *If $m_{e,j} < \infty$ for at least one $j \in S$, then*

$$D = \sum_{n\geqslant 0}(I+Q-\Pi)^n - \Pi.$$

Note that $m_{e,j}$ is finite for finite state space S. Hence, the above lemma provides a proof for the fact that for any finite Markov process the deviation matrix can be obtained through the inverse of $\Pi - Q$.

EXAMPLE 3.6. Consider a stable M/M/∞ queue. Then, $m_{e,0} < \infty$ and the deviation matrix exists and can be computed as described in Lemma 3.5.

We will frequently use results from Heidergott et al. (2007) on v-norms. Although the analysis in Heidergott et al. (2007) has been carried out for finite Markov chains, part of the results in Heidergott et al. (2007) are obtained by purely norm-theoretic arguments applied to products of matrices and can be carried over to the denumerable state-space case without any harm. In the following we give for easy reference a summary of the results that we will use from Heidergott et al. (2007).

Let A, B, C, F be square matrices (possibly infinite dimensional), and set

$$\mathrm{H}(n) = F\sum_{k=0}^{n}((A-B)C)^k,$$

$$\mathrm{R}(n) = F\sum_{k=n+1}^{\infty}((A-B)C)^k,$$

and

$$G = F\sum_{k=0}^{\infty}((A-B)C)^k,$$

provided that the sum converges, i.e., $G = H(n) + R(n)$ for all n. The following type of condition will be frequently used:

[A, B, C] *There exists a finite number N such that we can find $\delta_N \in (0,1)$, which satisfies:*

$$\left\|((A-B)C)^N\right\|_v < \delta_N,$$

and we set

$$c^v_{A,B,C} \stackrel{\text{def}}{=} \frac{1}{1-\delta_N}\left\|\sum_{k=0}^{N-1}((\mathrm{A}-\mathrm{B})\mathrm{C})^k\right\|_v.$$

The following result has been established in Heidergott et al. (2007).

LEMMA 3.7. *Suppose that $\|F\|_v$ is finite and consider the following statements*:

(i) *Condition* [A, B, C] *holds.*

(ii) *There exists $\rho \in (0,\infty)$ and $\delta \in (0,1)$ such that for all k,*

$$\|((A-B)C)^k\|_v \leqslant \rho\delta^k.$$

(iii) *For all k it holds that*

$$\|\mathrm{R}(k-1)\|_v \leqslant c^v_{A,B,C}\|F((\mathrm{A}-\mathrm{B})\mathrm{C})^k\|_v \quad \textit{and}$$

$$\|\mathrm{R}(k-1)\|_v \leqslant c^v_{A,B,C}\|F\|_v\rho\delta^k,$$

with δ and ρ as in (ii) *and $c^v_{A,B,C}$ as in* (i).

(iv) *$\|\mathrm{H}(n)\|_v$ converges to $\|G\|$ as n tends to infinity, and $\|\mathrm{R}(n)\|_v$ converges to zero as n tends to infinity at a geometric rate.*

(v) *$\mathrm{H}(n)$ converges to G as n tends to infinity.*

Then,

$$\text{(i)} \Rightarrow \text{(ii)} \Rightarrow \text{(iii)} \Rightarrow \text{(iv)} \Rightarrow \text{(v)}.$$

For finite matrices such that F has identical rows, it holds for all k that

$$\|\mathrm{R}(k-1)\|_v \leqslant c^{\mathbf{1}}_{A,B,C}\|F((A-B)C)^k\|_{\mathbf{1}},$$

where **1** *denotes the unit vector.*

Note that $\|F\|_v < \infty$ in the finite-dimensional case. In Heidergott et al. (2007), it has been shown that for finite matrices it holds that (i) to (v) in Lemma 3.7 are equivalent.

4. Series Representation for Denumerable Markov Processes

This section presents series expansions for continuous-time Markov processes with denumerable state space. The overall formula is derived in §4.1. In §4.2, sufficient conditions for the convergence of the series are presented. An application to sensitivity analysis is provided in §4.3. Eventually, we will present some new results on bounds on perturbations of Markov processes in §4.4.

4.1. The Update Formula

Let us now consider a Markov process $\mathscr{X}^*$ uniquely determined by its infinitesimal generator Q^*. To compute the corresponding ergodic matrix Π^*, we derive its series expansion based on the generator Q, the ergodic matrix Π, and the deviation matrix D of a process for which these are well known or easily calculated. By adding Q^*D to the equation put forward in Lemma 3.1(iv), we get

$$(Q^* - Q)D = Q^*D + I - \Pi.$$

Multiplying this equation with Π^* yields

$$\Pi^*(Q^* - Q)D = \Pi^* Q^* D + \Pi^* - \Pi^*\Pi,$$

which can be simplified by applying Lemma 3.1(i) and (ii) so that we obtain

$$\Pi^*(Q^* - Q)D = \Pi^* - \Pi.$$

This can be written as

$$\Pi^* = \Pi + \Pi^*(Q^* - Q)D. \tag{9}$$

Inserting (9) into its right side yields

$$\Pi^* = \Pi + \Pi(Q^* - Q)D + \Pi^*((Q^* - Q)D)^2. \tag{10}$$

Now, from inserting (9) into the right side of (10), we get

$$\Pi^* = \Pi + \Pi(Q^* - Q)D + \Pi((Q^* - Q)D)^2 + \Pi^*((Q^* - Q)D)^3.$$

By repeating this step n times we obtain

$$\Pi^* = \Pi \sum_{k=0}^{n} ((Q^* - Q)D)^k + \Pi^*((Q^* - Q)D)^{n+1}. \tag{11}$$

We call the representation in (11) the *continuous-time update formula*. The analogous formula for discrete-time Markov chains on a finite state space first appeared in Schweitzer (1968). By virtue of Lemma 3.1, the adoption of Schweitzer's result to continuous-time processes with denumerable state space follows reasonably from the line of arguments in Schweitzer (1968). We separate (11) into the *series approximation of degree n*

$$H(n) \stackrel{\text{def}}{=} \Pi \sum_{k=0}^{n} ((Q^* - Q)D)^k$$

and *remainder term*

$$R(n) \stackrel{\text{def}}{=} \Pi^*((Q^* - Q)D)^{n+1}.$$

The representation in (11) provides a scheme for approximately computing Π^*, provided that $R(n)$ tends to 0 as n tends to infinity. A sufficient condition for this is that $\|((Q^* - Q)D)^n\|_v$ tends to zero as n tends to infinity. Moreover, provided this condition holds, then for any f, such that $f(x) \leqslant cv(x)$ for all $x \in S$ and some c (in formula: $\|f\|_v < \infty$), it holds that $|\pi^* f - H(n)f|$ converges to zero as n tends to infinity.

A general sufficient condition for $\|R(n)\|_v$ to converge to zero is that D exists and that Q^* and Q are so close in the v-norm sense that $\|(Q^* - Q)D\|_v = \rho < 1$. For example, according to Example 3.6, the deviation matrix for the M/M/∞ system exists and choosing Q^* close to Q will guarantee convergence of $\|R(n)\|_v$ to zero. This idea has been used in Schweitzer (1968) in the context of perturbation analysis and in Kartashov (1996) in the context of stability analysis. Unfortunately, in applications the Q^* of interest typically fails to be close to Q in the above sense.

When we use the update formula for numerical purposes in §5, we will assume that a finite number N and constant $\delta_N < 1$ exist such that

$$\|((Q^* - Q)D)^N\|_v < \delta_N. \tag{12}$$

As we will show, condition (12) implies that $\|\Pi^* - H(n)\|_v$ tends to zero at an exponential rate.

Using simple norm arguments, it holds that

$$|\pi^* f - H(n)f| \leqslant \|\Pi^* - H(n)\|_v \|f\|_v \inf_s v(s).$$

Assuming that $v(s) = 1$ for at least one $s \in S$, we obtain

$$|\pi^* f - H(n)f| \leqslant \|\Pi^* - H(n)\|_v \|f\|_v. \tag{13}$$

Hence, provided that $\|\Pi^* - H(n)\|_v$ tends to zero at an exponential rate, it follows that $|\pi^* f - H(n)f|$ tends to zero as n tends to infinity at an exponential rate for any f with $\|f\|_v < \infty$.

Remark 4.1. Provided that Q and Q^* are uniformizable with rate μ and μ^*, respectively, the effect of switching from Q to Q^* on the stationary distribution can alternatively be expressed via the corresponding sampled chains. Let $\eta \geqslant \max(\mu, \mu^*)$. Recall that the stationary distributions of the sampled chain and the continuous-time process coincide: $\Pi_\eta = \Pi$ and $\Pi^*_\eta = \Pi^*$. Inserting this together with (6) into (11) yields the following alternative expansion:

$$\Pi^* = \Pi \sum_{k=0}^{n} ((P^*_\eta - P_\eta)D_\eta)^k + \Pi^*((P^*_\eta - P_\eta)D_\eta)^{n+1}. \tag{14}$$

We call the presentation in (14) the *sampled update formula*, and we denote by

$$H_\eta(n) \stackrel{\text{def}}{=} \Pi \sum_{k=0}^{n} ((P_\eta^* - P_\eta) D_\eta)^k$$

the *sampled series approximation of degree n* and by

$$R_\eta(n) \stackrel{\text{def}}{=} \Pi^* ((P_{\eta^*}^* - P_\eta) D_\eta)^{n+1}$$

the *remainder term*. Wheras the continuous-time update formula holds for continuous-time Markov processes with conservative generator matrix, the sampled update formula only applies to uniformizable processes, which excludes, for example, the M/M/∞ system.

In order to be able to apply the update formula in (11), one requires sufficient conditions such that $R(n)$ tends to zero as n tends to infinity. Such conditions are hard to get in the general case (i.e., for $R(n)$ without uniformization). A detailed discussion of sufficient conditions for convergence of $H(n)$ towards Π^* will be presented in §4.2.

4.2. Convergence of $H(n)$

For practical purposes it is important that convergence of $H(n)$ towards Π^* occurs at a geometric rate. Because the geometric rate with which the series converges is only computable in special cases, we propose to find the convergence rate in an iterative way. The key observation is that if there exists N such that $\|((Q^* - Q)D)^N\|_v < \delta_N$ for $\delta_N < 1$, then $H(k)$ converges at an exponential rate of at least δ_N. We introduce the following condition.

(C) *There exists a finite number N such that we can find* $\delta_N \in (0, 1)$*, which satisfies:*

$$\|((Q^* - Q)D)^N\|_v < \delta_N,$$

and we set

$$c^v_{\delta_N} \stackrel{\text{def}}{=} \frac{1}{1-\delta_N} \left\| \sum_{k=0}^{N-1} ((Q^* - Q)D)^k \right\|_v.$$

Note that (C) is in fact condition $[A, B, C]$ for $A = Q^*$, $B = Q$, and $C = D$. The factor $c^v_{\delta_N}$ in condition (C) allows us to establish an upper bound for the remainder term that is independent of Π^*.

Denote by $T(k) = \Pi((Q^* - Q)D)^k$ the kth element of the series in (11). Applying Lemma 3.7 for $A = Q^*$, $B = Q$, $C = D$, $F = \Pi$, and $G = \Pi^*$ yields the following result for the continuous-time update formula.

Lemma 4.2. *Under* (C) *it holds that for any* $v \geqslant 1$:

(i) $\|R(k-1)\|_v \leqslant c^v_{\delta_N} \|T(k)\|_v$ *for any k.*

(ii) $\lim_{k\to\infty} H(k) = \Pi \sum_{n=0}^{\infty} ((Q^* - Q)D)^n = \Pi^*$.

(iii) $\rho \in \mathbb{R}$ *and* $\delta < 1$ *exist such that* $\|((Q^* - Q)D)^k\|_v < \rho\delta^k$ *for all k.*

(iv) *For all k it holds that* $\|T(k)\|_v < \rho\delta^k \|\Pi\|_v$*, with* ρ *and* δ *as in* (iii).

Condition (C) is of key importance for our approach and this gives rise to the question of for which class of systems (C) holds. It has been shown in Heidergott et al. (2007) for finite discrete-time Markov chains that (the discrete-time counterpart of condition) (C) is equivalent to the convergence of $H(n)$ as n tends to infinity. Unfortunately, (C) is a stronger condition than v-norm ergodicity; see the example of a finite discrete-time Markov chain that is v-norm ergodic but fails to satisfy (C) in Heidergott et al. (2007). The counterexample in Heidergott et al. (2007) is, however, a rather esoteric Markov chain and in applications we have so far encountered no system that violates (C). In the following we will show that v-normed ergodicity of the sampled chain is sufficient for convergence of $H_\eta(n)$ as n tends to infinity applied to appropriate powers of the sampled chain. This result will hold without any condition of the type of (C).

Let Q^* and Q be uniformizable with rate μ^* and μ, respectively, and let $\eta = \max(\mu^*, \mu)$. In the following we will establish for uniformizable processes sufficient conditions for the convergence of the sum on the right-hand side in (14) for appropriate powers of P_η^* and P_η. The key condition is the following:

(C′) *There exist finite numbers k and N such that we can find* $\delta_N \in (0, 1)$*, which satisfies:*

$$\left\| \left(((P_\eta^*)^k - P_\eta^k) D_{\eta,k} \right)^N \right\|_v < \delta_N,$$

where $D_{\eta,k}$ *is the deviation matrix associated with* P_η^k*, and set*

$$c^v_{\delta_{N,\eta}} \stackrel{\text{def}}{=} \frac{1}{1-\delta_N} \left\| \sum_{n=0}^{N-1} (((P_\eta^*)^k - P_\eta^k) D_{\eta,k})^n \right\|_v.$$

Note that (C′) is in fact condition $[A, B, C]$ for $A = (P_\eta^*)^k$, $B = P_\eta^k$, and $C = D_{\eta,k}$.

As we will show in the next theorem, v-normed ergodicity of P_η^* and P_η implies that condition (C′) is satisfied for $k \geqslant 2$, and thereby yields convergence of $H_\eta(n)$ as n tends to infinity.

Theorem 4.3. *Let Q and* Q^* *be uniformizable, such that* P_η^* *and* P_η *are v-norm ergodic. Then, condition* (C′) *is satisfied for* $N \geqslant 2$ *and k sufficiently large, and it holds that*

$$\Pi^* = \Pi \sum_{n\geqslant 0} \left(((P_\eta^*)^k - P_\eta^k) D_{\eta,k} \right)^n$$

for any $k \geqslant 2$.

To summarize, (C) is a sufficient condition for $H(n)$ to converge as n tends to infinity and to guarantee that the v-norm of the remainder decreases at a geometric rate. On the other hand, without condition (C) the result of the next subsection holds. Uniformizability, together with v-normed ergodicity of the sampled chains, is sufficient for the sampled series $H_\eta(n)$ to converge as n tends to infinity and for the remainder term to decrease at geometric rate, provided the series is developed for powers of the sampled chains.

4.3. Sensitivity Analysis

Suppose that Q depends on some parameter $\theta \in (a, b) \subset \mathbb{R}$. For example, let Q be the infinitesimal generator of the continuous-time queue-length process of the M/M/1 queue with arrival rate λ and service rate μ; then, Q may be interpreted as a mapping of $\theta = \mu$ with $\theta \in (\lambda, \infty)$, in writing Q_θ. In perturbation analysis one is typically interested in the effect of a change in θ on the stationary distribution. More formally, let π_θ denote the stationary distribution associated with Q_θ; then, perturbation analysis seeks to compute $d\pi_\theta/d\theta$. The often simple structure of Q_θ motivates the following Lipschitz condition:

$$\text{(K)}\quad \forall\, i, j \in S\colon \quad \frac{1}{\Delta}\left|Q_{\theta+\Delta}(i, j) - Q_\theta(i, j)\right| \leqslant K.$$

Indeed, in the M/M/1 example the entries of Q_θ are linear mappings of θ. For given D_θ, let d_θ denote the vector of absolute column sums for D_θ, i.e.,

$$d_\theta(j) = \sum_i |D_\theta(i, j)|, \quad j \in S. \tag{15}$$

Theorem 4.4. *Let condition* (K) *be satisfied. If the vector d_θ defined in* (15) *is finite, then π_θ is continuous at θ. If in addition Q_θ is elementwise differentiable, then*

$$\pi'_\theta = \pi_\theta Q'_\theta D_\theta.$$

Finiteness of d_θ is guaranteed for finite Markov processes. In the denumerable case, a sufficient condition for d_θ to be finite is $\|D_\theta^T\|_{\mathbf{1}} < \infty$ where D_θ^T denotes the transpose of D_θ. It is worth noting that Theorem 4.4 applies without uniformization. To see this, note that the deviation matrix of M/M/∞ satisfies the condition put forward in Theorem 4.4, although the system fails to be uniformizable. In addition, Theorem 4.4 does not require Q to be strongly ergodic.[1] This is of importance because even such simple systems as the M/M/1 queue are v-norm ergodic for appropriate v, but fail to be strongly ergodic; see Hordijk and Spieksma (1992). Hence, the above perturbation formula extends the result in Cao and Chen (1997) to nonuniformizable and non-strongly ergodic processes.

4.4. Perturbation Bounds

The study of perturbation bounds for Markov chains is known as *perturbation analysis of Markov chains* (PAMC) in the literature. PAMC is a classical topic in Markov chain literature and dates back to Schweitzer (1968). The key task in PAMC is the following: Provide bounds on the effect of perturbing P to P^* on the stationary behavior. The above problem can be phrased as follows: Can $\|\pi^* - \pi\|_v$ be approximated or bounded in terms of $\|P^* - P\|_v$?

In the following, we provide a simple bound for denumerable continuous-time Markov processes. By (9) it holds that

$$\Pi^* = \Pi + \Pi^*(Q^* - Q)D.$$

Subtracting Π on both sides and taking v-norms yields

$$\|\Pi^* - \Pi\|_v = \|\Pi^*(Q^* - Q)D\|_v,$$

which yields

$$\|\Pi^* - \Pi\|_v \leqslant \|\Pi^*\|_v \|Q^* - Q\|_v \|D\|_v.$$

Provided that condition (C) holds, it follows from Lemma 4.2 that $\|\Pi^*\|_v \leqslant c^v_{\delta_N} \|\Pi\|_v$, and we arrive at

$$\|\Pi^* - \Pi\|_v \leqslant c^v_{\delta_N} \|\Pi\|_v \|Q^* - Q\|_v \|D\|_v. \tag{16}$$

The following lemma lists some bounds on perturbations.

Lemma 4.5.
(i) *Provided that* (C) *holds, we have*

$$\|\Pi^* - \Pi\|_v \leqslant c^v_{\delta_N} \|\Pi\|_v \|Q^* - Q\|_v \|D\|_v.$$

(ii) *Let Q be uniformizable with rate μ. Suppose that c and β, with $0 \leqslant \beta < 1$, exist such that $\|(P_\mu)^n - \Pi\|_v \leqslant c\beta^n$ for all n* (*in words, P_μ is v-norm ergodic*). *If* (C) *holds, then*

$$\|\Pi^* - \Pi\|_v \leqslant c^v_{\delta_N} \|\Pi\|_v \|Q^* - Q\|_v \frac{c}{\mu(1-\beta)}.$$

(iii) *Let Q and Q^* be uniformizable with rate η. Suppose that c and β, with $0 \leqslant \beta < 1$, exist such that $\|(P_\mu)^n - \Pi\|_v \leqslant c\beta^n$ for all n* (*in words, P_μ is v-norm ergodic*). *Then constants c_k exist such that for each $k \geqslant 2$ it holds that*

$$\|\Pi^* - \Pi\|_v \leqslant c_k \|\Pi\|_v \|(P^*_\eta)^k - P^k_\eta\|_v \frac{c}{\eta(1-\beta^k)}.$$

To the best of our knowledge, Lemma 4.5 is a first result on a perturbation bound for a denumerable continuous-time Markov process. The statement in Lemma 4.5 is also of interest for the study of strong stable Markov processes (for details, see Kartashov 1996) because it states that any uniformizable continuous-time Markov process with v-norm ergodic sampled chain is strongly stable provided that condition (C) holds.

5. Numerical Algorithm for Finite Markov Processes

In the following we will show how the continuous-time update formula can be made fruitful for numerical approximations. The numerical algorithm is presented in §5.1. The performance of the algorithm is illustrated in §5.2 with numerical examples.

5.1. The Algorithm

With Lemma 4.2 we arrive at the following numerical approach. First, we search for N such that $1 > \delta_N \stackrel{\text{def}}{=} \|((Q^* - Q)D)^N\|_v$. In words, we establish the minimal power of $((Q^* - Q)D)$ that yields geometrical convergence of $H(n)$. Then, we choose a precision ϵ up to which we want to approximate Π^*. The algorithm computes the elements $T(k)$ of $H(k)$ until our upper bound for $R(k)$, given by $c^v_{\delta_N}\|((Q^* - Q)D)^{k+1}\|_v$, drops below ϵ.

We can now describe an algorithm that yields an approximation for π^* with ϵ precision.

Algorithm 1

Choose precision $\epsilon > 0$. Set $k = 1$, $T(1) = \Pi(Q^* - Q)D$, and $H(0) = \Pi$.

Step 1. Find N such that $\|((Q^* - Q)D)^N\|_v < 1$. Set $\delta_N = \|((Q^* - Q)D)^N\|_v$ and compute

$$c^v_{\delta_N} = \frac{1}{1 - \delta_N}\left\|\sum_{k=0}^{N-1}((Q^* - Q)D)^k\right\|_v.$$

Step 2. If $c^v_{\delta_N}\|T(k)\|_v < \epsilon$, the algorithm terminates and $H(k-1)$ yields the desired approximation. Otherwise, go to step 3.

Step 3. Set $H(k) = H(k-1) + T(k)$. Set $k := k+1$ and $T(k) = T(k-1)(Q^* - Q)D$. Go to step 2.

The norm approximation provided by the algorithm implies approximation for expected costs. To see this, note that

$$|\pi^* f - H(k)f| \leqslant \|\Pi^* - H(k)\|_v \|f\|_v \inf_s v(s)$$

for any performance function f with $\|f\|_v < \infty$; see (13). Hence, if Algorithm 1 terminates after $\hat{k}$ iterations, then this yields for the specific performance function f the approximation

$$|\pi^* f| \approx \|H(\hat{k})\|_v \|f\|_v \inf_s v(s), \tag{17}$$

where the error of the above approximation is bounded by $\epsilon\|f\|_v \inf_s v(s)$.

If condition (C) holds, then the above algorithm terminates geometrically fast; see Lemma 4.2. In case S is finite, condition (C) and (ii) in Lemma 4.2 are equivalent, which allows one to empirically check whether condition (C) is satisfied. Moreover, for finite S all norms are equivalent with respect to norm ergodicity, and without loss of generality, we take $v = \mathbf{1}$, with $\mathbf{1}$ the vector with all elements equal to one, for the algorithm. Note that for this choice of v, $\inf_s v(s) = 1$ in (17).

5.2. Example: Approximating the Retrial Queue Through Simple Queues

We illustrate the use of Algorithm 1 for computing the ergodic distribution of a queueing process with impatient customers who recall after they hung up. The optimization of such a system was thus far hindered by the fact that a representation of the stationary distribution was not available in general. However, by applying the series approximation $\lim_{n\to\infty} H(n)$ introduced in §4, we can approximate Π^*.

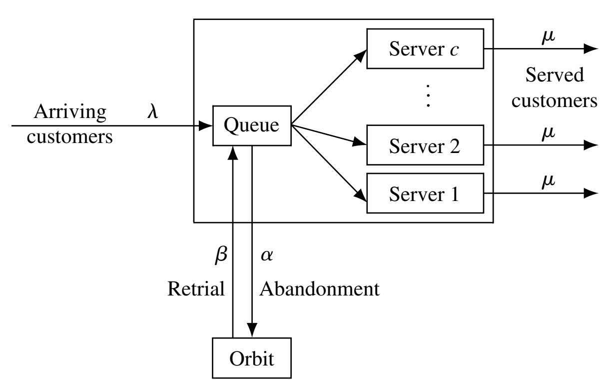
Let $\mathscr{X}^*$ be the ergodic queue-length process with states $(x_1, x_2)^t \in S \stackrel{\text{def}}{=} \mathbb{N}_0 \times \mathbb{N}_0$, where x_1 denotes the number of customers either in service or waiting in the queue and x_2 refers to the impatient customers intending to recall. We regard this model as an open Jackson network with two nodes. External arrivals—modeling first callers—enter the system with rate λ at the first node, where they are served by c servers each providing service at rate μ. However, callers abandon if their waiting time exceeds their exponentially-α distributed patience. Customers who hung up are considered to enter a second node—the orbit—which they leave by recalling after an exponentially-β distributed time. Therefore, the first node is an M/M/c queue with abandonments, whereas the latter one is an M/M/∞ queue. Transition rates for x, $y \in S$ are given as follows:

$$q^*_{x,y} = \begin{cases} \lambda & y=(x_1+1,x_2),\ x_1,x_2 \geqslant 0 \\ \min\{x_1,c\}\mu & y=(x_1-1,x_2),\ x_1 \geqslant 1,\ x_2 \geqslant 0 \\ \max\{x_1-c,0\}\alpha & \\ & y=(x_1-1,x_2+1),\ x_1 \geqslant 1,\ x_2 \geqslant 0 \\ x_2\beta & y=(x_1+1,x_2-1),\ x_1 \geqslant 0,\ x_2 \geqslant 1 \\ -(\lambda+\min\{x_1,c\}\mu+\max\{x_1-c,0\}\alpha+x_2\beta) & \\ & y=(x_1,x_2),\ x_1,x_2 \geqslant 0 \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

An overview of the system is provided in Figure 1.

Retrial queues have been intensively studied in the literature; see, for example, Falin and Templeton (1997), Kulkarni and Liang (1997), and the references therein, as

Figure 1. Structure of an M/M/c + M queueing system with abandonment and retrial.



well as the two survey papers (Yang and Templeton 1987 and Falin 1990). Retrial queues are especially useful for modelling call centers; see, for example, Gans et al. (2003), but also have some further applications, of which four are presented in Yang and Templeton (1987). For a recent overview on retrial queues, we refer to Artalejo (2008). A closed-form solution for the stationary distribution of the M/M/c retrial queue is only available for $c = 1, 2$; see Falin and Templeton (1997). For larger values of c, only approximations are known (Artalejo 2008). In the following we will approximate the M/M/c retrial queue for general c. To this end, we introduce a related process $\mathscr{X}$, with x_1 the number of customers waiting or being served in an M/M/c queue and x_2 the number of customers being served in an independently acting M/M/∞ queue. The arrival and service rates of the first queue remain the same as in the initial model, whereas external customers enter the M/M/∞ queue at rate α and leave the system after being served at rate $x_2\beta$; see Figure 2.

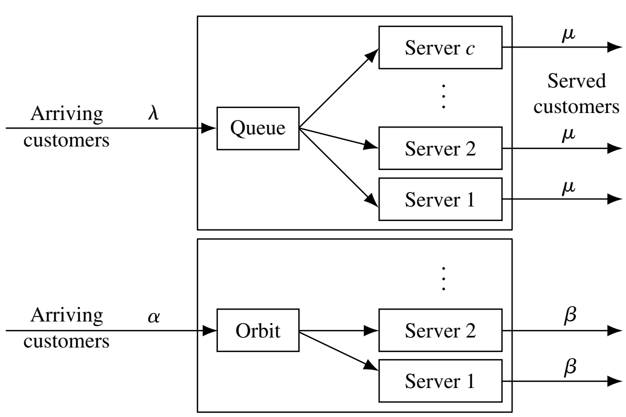
Therefore it holds for the transition rates with $x, y \in S$

$$q_{x,y} = \begin{cases} \lambda & y=(x_1+1, x_2),\, x_1, x_2 \geqslant 0 \\ \min\{x_1, c\}\mu & y=(x_1-1, x_2),\, x_1 \geqslant 1, x_2 \geqslant 0 \\ \alpha & y=(x_1, x_2+1),\, x_1, x_2 \geqslant 0 \\ x_2\beta & y=(x_1, x_2-1),\, x_1 \geqslant 0, x_2 \geqslant 1 \\ -(\lambda+\min\{x_1, c\}\mu+\alpha+x_2\beta) & \\ & y=(x_1, x_2),\, x_1, x_2 \geqslant 0 \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

and we have the joint ergodic distribution

$$\pi_x = \left(\sum_{k=0}^{\infty} \frac{c^{\min\{c-k,0\}}(\lambda/\mu)^k}{\min\{k, c\}!}\right)^{-1} \cdot \frac{c^{\min\{c-x_1,0\}}(\lambda/\mu)^{x_1}}{\min\{x_1, c\}!} e^{-\alpha/\beta} \frac{(\alpha/\beta)^{x_2}}{x_2!}, \quad x \in S. \tag{20}$$

Figure 2. Structure of an M/M/c and M/M/∞ queueing system.



To receive the stationary distribution of our initial model we have to compute the remaining parts of $H(n)$ first. From (18) and (19) we get the entries of $(Q^* - Q)$ as follows:

$$(Q^*-Q)_{x,y} = \begin{cases} -\alpha & y=(x_1, x_2+1),\, x_1, x_2 \geqslant 0 \\ \max\{x_1-c, 0\}\alpha & y=(x_1-1, x_2+1),\, x_1 \geqslant 1, x_2 \geqslant 0 \\ -x_2\beta & y=(x_1, x_2-1),\, x_1 \geqslant 0, x_2 \geqslant 1 \\ x_2\beta & y=(x_1+1, x_2-1),\, x_1 \geqslant 0, x_2 \geqslant 1 \\ \alpha-\max\{x_1-c, 0\}\alpha & y=(x_1, x_2),\, x_1, x_2 \geqslant 0 \\ 0 & \text{otherwise.} \end{cases}$$

Now we apply Algorithm 1 to a retrial queue with arrival rate of external arrivals $\lambda = 3$ and $c = 3$ servers, each providing service at rate $\mu = 1$. Customers abandon with rate $\alpha = 0.5$, and customers who abandoned leave the orbit in order to recall with rate $\beta = 3$. The service center has a capacity of $N_1 = 4$ and the orbit one of $N_2 = 6$. Whenever an arriving customer to the service center or the orbit finds no available space in the respective queue, he is lost. We will focus our attention on the mean stationary number of customers at the service center, i.e., customers in orbit are not counted. More formally, let $w(x, y) = x$ for $x, y \in S$, our goal is to approximately compute $\pi^* w$.

Figure 3(a) shows the absolute error in percentage for predicting $\pi^* w$ by $H(n)w$, i.e., the figure plots $|\pi^* w - H(n)w|/\pi^* w$ as a mapping of n. The numerical value of $\pi^* w$ is 2.4947.

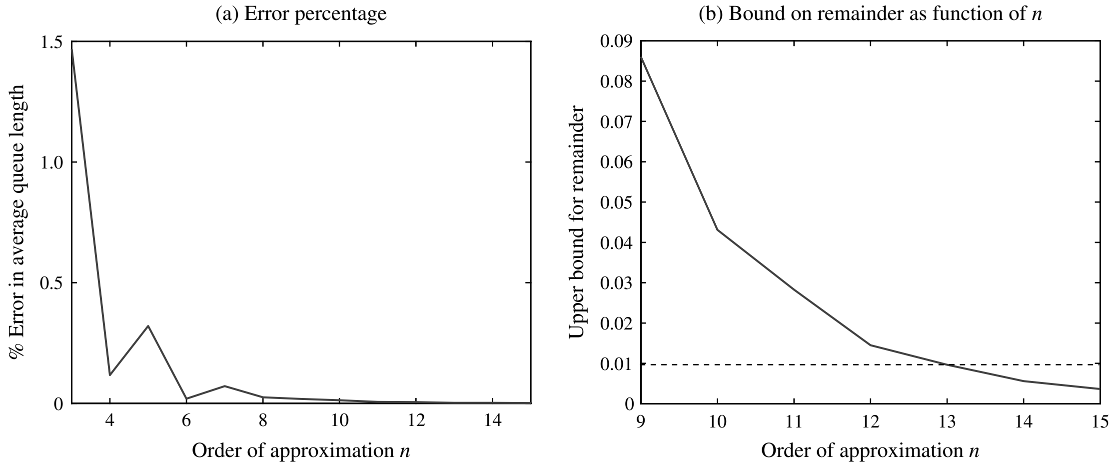
Algorithm 1 terminates whenever $c^v_{\delta_N} \|T(n)\|_v$ drops below ϵ. For example, taking $\epsilon = 0.01$, the algorithm will compute $\pi^* w$ up to a precision of $\pm 0.01\|w\|_{\mathbf{1}}$, where $\|w\|_{\mathbf{1}} = 4$. The number of elements of $H(n)$ required for achieving this precision is illustrated in Figure 3(b), where the dotted line represents ϵ. As can be seen from Figure 3(b), $H(13)$ yields the desired precision, and increasing the precision to, say, 0.005 would lead to $H(15)$. For the sake of completeness we state the values required in Algorithm 1. For the above example, we obtained $N = 11$, $\delta_N = 0.9179$, and $c^{\mathbf{1}}_{\delta_N} = 201.2311$.

As we have mentioned earlier, for finite state Markov processes, Algorithm 1 can also be formulated for the sampled chain using the sampled update formula. The algorithmic complexity of both algorithms will be identical, because for both algorithms the main numerical work is in computing Π and D.

6. Functional Series Expansion

In this section we present a functional version of the series expansion that allows one to obtain Π as a function of θ, where θ is a parameter of Q. The functional model is introduced in §6.1. The application to performance evaluation of

Figure 3. Predicting the stationary queue length: Error percentage (left) and bound on remainder term (right).



finite state Markov processes is presented in §6.2. Eventually, we illustrate the performance of the algorithm in §6.3 with numerical examples.

6.1. The Basic Model

As has already been noted in §4.3, the elements of Q are typically linear mappings of the rates of the Markov process. Consider two generator matrices Q^* and Q, and define

$$Q_\theta = \theta Q^* + (1-\theta)Q, \quad \theta \in [0, 1]. \tag{21}$$

Denote the stationary distribution associated with Q_θ by Π_θ. The basic model implies that

$$Q_{\Delta+\theta} - Q_\theta = \Delta(Q^* - Q),$$

for any Δ such that $\theta + \Delta \in [0, 1]$. Inserting the above representation for $Q_{\Delta+\theta} - Q_\theta$ into (11) yields

$$\Pi_{\Delta+\theta} = \Pi_\theta \sum_{k=0}^{n} \Delta^k ((Q^* - Q)D_\theta)^k + \Delta^{n+1} \Pi_{\Delta+\theta} ((Q^* - Q)D_\theta)^{n+1},$$

with D_θ the deviation matrix associated with Q_θ. Provided that the remainder term on the right-hand side of the above formula tends to zero as n tends to infinity, one obtains

$$\Pi_{\Delta+\theta} = \Pi_\theta \sum_{k=0}^{\infty} \Delta^k ((Q^* - Q)D_\theta)^k, \tag{22}$$

which is noticeably the Taylor series expansion of Π_θ developed at point θ. Taylor series expansions of Markov chains are a topic of active research. A Taylor series expansion for irreducible discrete-time Markov chains on denumerable state space can be found in Cao (1998). This result has been extended to v-norm ergodic discrete-time Markov chains on general state space in Heidergott and Hordijk (2003). For discrete-time Markov chains on denumerable state space with several ergodic classes, a Taylor series expansion can be found in Altman et al. (2004).

We call the presentation in (22) the *continuous-time update formula*. We separate (22) into the *series approximation of degree n*

$$H_\Delta(n) \stackrel{\text{def}}{=} \Pi_\theta \sum_{k=0}^{n} \Delta^k ((Q^* - Q)D_\theta)^k$$

and *remainder term*

$$R_\Delta(n) \stackrel{\text{def}}{=} \Delta^{n+1} \Pi_{\Delta+\theta} ((Q^* - Q)D_\theta)^{n+1}.$$

6.2. The Algorithm

Following the line of thought put forward in §5, we obtain the following algorithm for approximating $\pi_{\Delta+\theta}$ with ϵ precision, where $T_\Delta(k) = \Delta^k \Pi_\theta ((Q^* - Q)D_\theta)^k$ denotes the kth element in $H_\Delta(n)$.

Algorithm 2
Choose precision $\epsilon > 0$.

Step 1. Find N such that $\|((Q^* - Q)D_\theta)^N\|_v < 1$. Set $\delta_N = \|((Q^* - Q)D_\theta)^N\|_v$ and compute

$$c^v_{\delta_N} = \frac{1}{1-\delta_N} \left\| \sum_{k=0}^{N-1} ((Q^* - Q)D_\theta)^k \right\|_v.$$

Step 2. For any Δ such that $|\Delta| < 1/\delta_N$, compute as follows. Set $k = 1$, $T_\Delta(1) = \Delta \Pi_\theta (Q^* - Q) D_\theta$, and $H_\Delta(0) = \Pi_\theta$.

Step 3. If $c^{\mathbf{1}}_{\delta_N} \|T_\Delta(k)\|_v < \epsilon$, the algorithm terminates and $H_\Delta(k-1)$ yields the desired approximation. Otherwise, go to step 4.

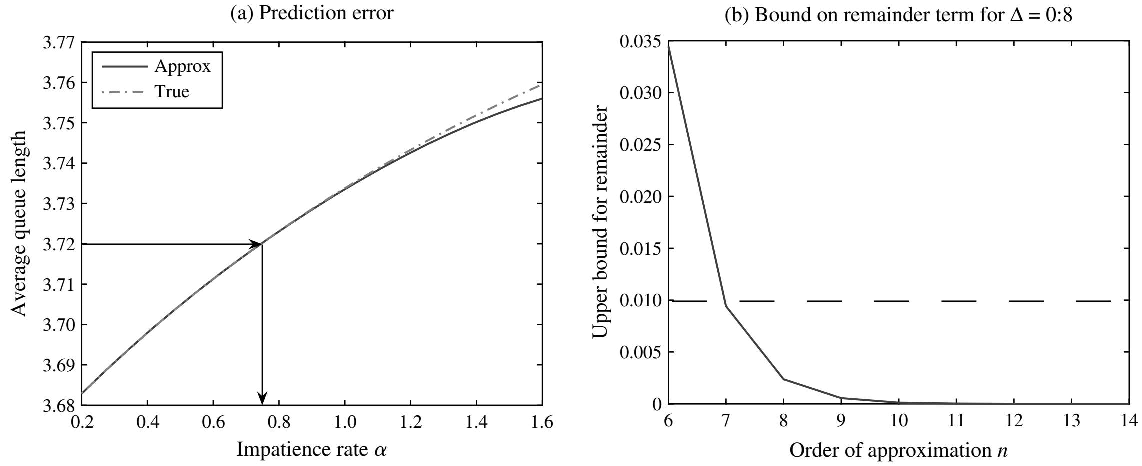
Step 4. Set

$$T_\Delta(k+1) = \Delta T_\Delta(k)(Q^* - Q)D_\theta$$

and $H_\Delta(k) = H_\Delta(k-1) + T_\Delta(k)$. Let $k := k+1$ and go to step 3.

Note that the convergence properties of the above algorithm are just like Algorithm 1. In particular, we take $v = \mathbf{1}$ in case S is finite.

Figure 4. Predicting the stationary queue length: Error percentage (left) and bound on remainder term (right).



6.3. Example: Dependence of the Retrial Queue on Impatience Rate

We illustrate Algorithm 2 with the following example. The arrival rate of external arrivals is $\lambda = 3$, and there are $c = 2$ servers, with each providing service at rate $\mu = 1$. Customers who abandoned leave the orbit in order to recall with rate $\beta = 3$. Furthermore, we assume that the service center has a total capacity of $N_1 = 5$ and orbit of $N_2 = 6$. Customers who find no available space upon arrival to the service center or the orbit are lost. The performance measure we are interested in is the mean stationary number of customers at the service center, i.e., customers in orbit are not counted. More formally, let $w(x, y) = x$ for $x, y \in S$, our goal is to approximately compute $\pi^* w$. Let Q^* be the generator matrix of the finite state-space version of the M/M/c system with abandonments and retrial with customers abandon with rate $\alpha^* = 0.9$, and let Q be the generator matrix of the same system except for the abandon rate, which is set to $\alpha = 0.2$.

The generator Q_α is an affine linear mapping in α, and we will in the following consider the parametric model

$$Q_{\alpha+\Delta} = Q_\alpha + \frac{\Delta}{\alpha^* - \alpha}(Q_{\alpha^*} - Q_\alpha), \quad 0 \leqslant \Delta \leqslant \alpha^* - \alpha.$$

Note that the above model differs from the model put forward in (21) only with respect to the range of Δ. In order to adjust Algorithm 2 to the above setting, replace Δ by $\Delta/(\alpha^* - \alpha)$. In particular, the termination condition of the algorithm in Step 2 then reads

$$\left|\frac{\Delta}{\alpha^* - \alpha}\right| < \frac{1}{\delta_N} \;\Leftrightarrow\; |\Delta| < \frac{\alpha^* - \alpha}{\delta_N}.$$

Figure 4(a) shows $\pi^* w$ and $H_{\alpha+\Delta}(n)w$ as a mapping of $\Delta \in [0, 1.4]$ for $n = 4$. The values of auxiliary variables in Algorithm 2 are $N = 4$, $\delta_N = 0.7469$, and $c^{\mathbf{1}}_{\delta_N} = 7.2953$. Hence, $H_{\alpha+\Delta}(n)w$ converges as n tends to ∞ to the true mean stationary queue length for $|\Delta| < (\alpha^* - \alpha)/\delta_N = 0.7/0.7469 = 0.9372$. The Taylor series is depicted in Figure 4(a).

Suppose that one is interested in the maximal abandon rate that will result in a stationary mean queue length of, say, 3.72. Then, solving $3.72 = H_{\alpha+\Delta} w$ yields $\Delta = 0.523$ and the maximal value for α is given by $\alpha' = 0.723$, see the arrows in Figure 4(a). Hence, α' solves the optimization problem $\max_\alpha \pi_\alpha w \leqslant 3.72$.

Figure 4(b) shows the decay in our bound on the remainder term as a mapping of n for $\Delta = 0.8$. The horizontal line indicates the precision value $\epsilon = 0.01$. The minimal number n for which $H_{\alpha+\Delta}(n)$ has to be evaluated in order to guarantee $\|\Pi_{\alpha+\Delta} - H_{\alpha+\Delta}(n)\|_{\mathbf{1}} < \epsilon$ can be found to be the first integer such that the bound on the remainder drops below the horizontal line. As Figure 4(b) illustrates, we have $n = 7$ for $\Delta = 0.8$. In the same vein, numerical experiments yield $n = 5$ for $\Delta = 0.4$, and $n = 11$ for $\Delta = 1.2$.

A lower bound for the radius of convergence of $H_{\alpha+\Delta}(n)$ is given by $(\alpha^* - \alpha)/\delta_N$. Hence, decreasing δ_N increases the range for Δ for which we can be sure that $H_{\alpha+\Delta}(n)$ converges to the correct value. This can be achieved by adjusting the computation of δ_N in Algorithm 2. For example, if a radius of convergence of at least, say, $r > 0$ is required, then Step 1 in Algorithm 1 has to be modified so that N is determined by $\|((Q^* - Q)D)^N\|_{\mathbf{1}} < 1/r$.

This version of the algorithm will terminate in finite time for any $r > 0$. Obviously, Step 1(b) will require a longer time than Step 1 because the initial phase of $\|((Q^* - Q)D)^n\|_{\mathbf{1}}$ before geometric decay occurs might be rather long. On the other hand, simply taking α^* larger increases $\|Q^* - Q\|_{\mathbf{1}}$ and will also decrease the performance of the algorithm. The analysis of the trade-off between increased radius of convergence and longer initial phase of the modified algorithm is a topic for further research.

7. Conclusion

We presented a technique for series expansions of continuous-time Markov processes with denumerable state space. Our approach is in particular applicable to nonuniformizable Markov models. The efficiency of the algorithm has been illustrated with numerical experiments of finite Markov processes. The algorithm is applicable to infinite Markov processes and future research will investigate possible applications to infinite models whose deviation matrix can be obtained in closed form. In this setting, the particular choice v in the v-norm becomes important, and further research on finding the best choice for v is needed.

8. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Endnote

1. A Markov chain is called strongly ergodic if it is v-norm ergodic for $v \equiv 1$.

References

Altman, E., K. Avrachenkov, R. Núñez-Queija. 2004. Perturbation analysis for denumerable Markov chains with application to queueing models. *Adv. Appl. Probab.* **36** 839–853.

Artalejo, J. 2008. *Retrial Queueing Systems*. Springer, New York.

Cao, X.-R. 1998. The Maclaurin series for performance functions of Markov chains. *Adv. Appl. Probab.* **30** 676–692.

Cao, X.-R., H.-F. Chen. 1997. Perturbation realization, potentials, sensitivity analysis of Markov processes. *IEEE Trans. Automatic Control* **42** 1382–1393.

Coolen-Schrijner, P., E. van Doorn. 2002. The deviation matrix of a continuous-time Markov chain. *Probab. Engrg. Informational Sci.* **16** 351–366.

Di Crescenzo, A. 1994a. On some transformations of bilateral birth-and-death processes with applications to first passage time evaluations. *SITA Proc. 17th Sympos. Inform. Theory Its Appl.*, Hiroshima, Japan, 739–742.

Di Crescenzo, A. 1994b. On certain transformation properties of birth-and-death processes. R. Trappl, ed. *Cybernetics and Systems*. World Scientific, Singapore, 839–846.

Falin, G. 1990. A survey of retrial queues. *Queueing Systems* **7** 127–168.

Falin, G., J. Templeton. 1997. *Retrial Queues*. Chapman and Hall, London.

Gans, N., G. Koole, A. Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing Service Oper. Management* **5** 79–141.

Heidergott, B., A. Hordijk. 2003. Taylor series expansions for stationary Markov chains. *Adv. Appl. Probab.* **35** 1046–1070.

Heidergott, B., A. Hordijk, M. van Uitert. 2007. Series expansions for finite-state Markov chains, *Probab. Engrg. Informational Sci.* **21** 381–400.

Hordijk, A., F. Spieksma. 1992. On ergodicity and recurrence properties of a Markov chain with an application to an open Jackson network. *Adv. Appl. Probab.* **24** 343–376.

Kartashov, N. 1996. *Strong Stable Markov Chains*. VSP, Utrecht, The Netherlands.

Kemeny, J., J. Snell. 1960. Potentials for denumerable Markov chains. *J. Math. Anal. Appl.* **3** 196–260.

Koole, G., A. Mandelbaum. 2002. Queueing models for call centers. *Ann. Oper. Res.* **113** 41–59.

Kulkarni, V., H. Liang. 1997. Retrial queues revisited. J. Dshalalow, ed. *Frontiers in Queueing: Models and Applications in Science and Engineering*. CRC Press, Boca Raton, FL, 19–34.

Lenin, R., P. Parthasarathy, W. Scheinhardt, E. van Doorn. 2000. Families of birth-and-death process with similar time-dependent behaviour. *J. Appl. Probab.* **37** 835–849.

Leskelä, L. 2006. Stabilization of an overloaded queueing network using measure-based admission control. *J. Appl. Probab.* **43** 231–244.

Mitrophanov, A. 2004. The spectral gap and perturbation bounds for reversible continuous-time Markov chains. *J. Appl. Probab.* **41** 1219–1222.

Mitrophanov, A. 2005a. Sensitivity and convergence of uniformly ergodic Markov chains. *J. Appl. Probab.* **42** 1003–1014.

Mitrophanov, A. 2005b. Ergodicity coefficient and perturbation bounds for continuous-time Markov chains. *Math. Inequalities Appl.* **8** 159–168.

Mitrophanov, A. 2006. Estimates of sensitivity to perturbations for finite homogeneous continuous-time Markov chains. *Theory Probab. Its Appl.* **50** 319–326.

Mitrophanov, A., M. Borodovsky. 2006. Statistical significance in biological sequence analysis. *Briefings in Bioinformatics* **7** 2–24.

Mitrophanov, A., A. Lomsadze, M. Borodovsky. 2005. Sensitivity of hidden Markov models. *J. Appl. Probab.* **42** 632–642.

Pollett, P. 2001. Similar Markov chains. *J. Appl. Probab.* **38A** 53–65.

Schweitzer, E. 1968. Perturbation theory and finite Markov chains. *J. Appl. Probab.* **5** 401–413.

Shardlow, T., A. Stuart. 2000. A perturbation theory for ergodic properties of Markov chains. *SIAM J. Numer. Anal.* **37** 1120–1137.

Takahashi, Y. 1973. On the effects of small deviations in the transition matrix. *J. Oper. Res. Soc. Japan* **16** 104–129.

Tweedie, R. 1980. Perturbations of countable Markov chains and processes. *Ann. Institute Statist. Math.* **32** 283–290.

Yang, T., J. Templeton. 1987. A survey on retrial queues. *Queueing Systems* **2** 201–233.

Yin, G., Q. Zhang, G. Badowski. 2002. Asymptotic properties of singularly perturbed Markov chain with inclusion of transient states. *Ann. Appl. Probab.* **10** 549–572.